In the name of Allah, the Most Gracious, the Most Merciful









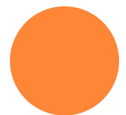
STATISTIQUE DESCRIPTIVE

Université d'Alger
Faculté de médecine
Département de médecine
Dr G. BRAHIMI – Dr A. AOUNI

OBJECTIF GÉNÉRAL

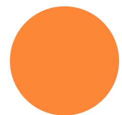
- Fournir à l'étudiant des notions de base en statistique descriptive pour le recueil, le traitement et l'interprétation des données recueillies en vue d'aider à la prise de décision dans le domaine de la santé.

- Apprécier l'importance de vérifier les méthodes de recueil et l'utilisation pratique de celles –ci.



OBJECTIFS SPÉCIFIQUES

- L'étudiant doit être capable de :
- Définir le type de variables
- Mettre en ordre les données d'un caractère qualitatif et quantitatif
- Traiter les données
- Présenter les données sous forme de tableaux
- Déterminer et apprécier l'intérêt des fréquences cumulées
- Définir, déterminer, calculer les paramètres de tendance centrale, moyenne, médiane, mode
- Définir, calculer les paramètres de dispersion variance, écart type.
- Apprécier les principaux avantages et inconvénients des paramètres de réduction
- Représenter graphiquement les données d'une variable qualitative
- Représenter graphiquement les données d'une variable quantitative
- Représenter graphiquement une distribution à 2 variables qualitatives.
- Calculer, interpréter
- Effectifs cumulés, fréquences relatives, fréquences relatives cumulées, moyenne, médiane, mode, variance, écart type, coefficient de variation.



DÉFINITIONS

- **G. Morlat : la statistique : "Le mot «statistique » désigne à la fois un ensemble de données, d'observation et l'activité qui consiste dans leur recueil, leur traitement et leur interprétation."**

- **Georges Morlat ajoute à ce sujet : " En réalité, ces deux sens du mot statistique ont naturellement entre eux des liens étroits. Il serait vain de recueillir des données, si ce n'était pour les traiter et les interpréter en vue d'éclairer les actions humaines ou de faire progresser la connaissance des phénomènes. Inversement, la manière de recueillir des données peut et doit être influencée d'abord par les méthodes de traitement ultérieures et ensuite par l'utilisation pratique de ces données ou des produits qui en sont dérivés.**

On distingue deux domaines :

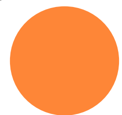
la statistique descriptive (statistique au sens commun du terme) appelé communément les statistiques.

Elle permet de recueillir, de mettre en forme les données sous forme de tableaux ou de graphiques, de synthétiser l'information recueillie, de la traiter et de l'interpréter afin d'en faciliter la connaissance.

la statistique inférentielle.

Elle permet, à partir d'un nombre réduit d'observations, d'étendre, de généraliser, sous certaines conditions, les conclusions obtenues à la population, d'estimer et de prédire.

La statistique permet de «dénombrer, calculer, mesurer, estimer, juger», et ainsi d'évaluer les phénomènes dans le domaine de la santé.



DÉFINITIONS

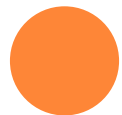
- **Population** (encore appelée ensemble)

C'est l'ensemble de tous les individus dont on cherche à déterminer une ou plusieurs caractéristiques, chaque individu est distinct des autres.

L'individu (encore appelée élément ou unité statistique) est l'entité élémentaire de base observée par le statisticien. Cela peut être une personne, un animal, un événement.

- **Echantillon :**
- C'est un sous ensemble de taille finie d'une population, une fraction de la population étudiée d'effectif volontairement réduit.
- C'est une collection d'éléments prélevés d'une façon particulière de la population statistique afin de tirer des conclusions sur cette population.
- Pour que les résultats puissent être généralisés à la population statistique, l'échantillon doit être représentatif de cette dernière, il doit refléter fidèlement sa composition et sa complexité, il est aléatoire (pris au hasard).

- **Série (ou distribution) :**
- C'est un ensemble de données (de valeurs) relatives à une variable mesurée sur un échantillon ou une population d'éléments. La valeur est la mesure.



DEFINITIONS

Modalité :

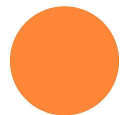
Une modalité est une catégorie de la variable étudiée ou une valeur qui affecte un caractère.

Exemple :

- Le sexe est une variable à deux modalités masculin et féminin, la maladie est une variable parce qu'on peut lui définir au moins deux modalités : malade et non malade, la glycémie est aussi une variable puisque ce caractère présente différentes valeurs.
- Il est primordial de bien définir les modalités des caractères de l'étude pour éviter les difficultés qui ne manqueront pas de surgir lors de la collecte et de l'exploitation.

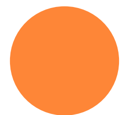
Ratio

- Un ratio est le rapport des fréquences (effectifs ou fréquences relatives) de deux modalités d'une même variable. Le numérateur n'est pas compris dans le dénominateur. Généralement le numérateur et le dénominateur se réfèrent à deux catégories d'une même variable.
- Exemple :
- sex-ratio = $\frac{\text{effectif masculin}}{\text{effectif féminin}}$



DEFINITIONS

- **Taux**
- Le taux mesure la probabilité de survenue d'un évènement donné au cours du temps. Un taux doit toujours s'exprimer en fonction d'une certaine unité de temps, pour un lieu géographique donné et pour un groupe de personnes bien défini.
- Le numérateur est un nombre d'évènements (décès, maladie, handicap) survenus au cours d'une certaine période t1 - t2. Le dénominateur représente la population exposée au risque de survenue de cet évènement pendant cette période.
- Exemple :
- Si dans les 24 heures qui suivent un repas pris dans une cantine fréquentée par 300 personnes, 30 présentent des signes d'intoxication alimentaire, le taux de la maladie est : 30/300 = 0.10 = 10.0 % = 100 pour mille.



DEFINITIONS

- **Indices élémentaires**
- Un indice est généralement le rapport entre deux mesures, c'est une pseudo - fréquence relative ; c'est un substitut d'une fréquence relative difficile à calculer. (Le numérateur n'est pas compris dans le dénominateur).
- L'indice est utilisé quand le dénominateur est difficile à déterminer.
- Exemple :
- L'indice de masse corporelle IMC se calcule en divisant le poids exprimé en kilogrammes par la taille au carré exprimée en mètre :
- **IMC = poids (kg) / taille au carré (exprimée en m^2)**.
- Les indices simples peuvent aussi mesurer l'évolution de 02 valeurs par une grandeur simple au cours de 02 périodes différentes.
- Indice =
- Gt est la valeur de la grandeur à la période " t"
- G1 est la valeur de la grandeur à la période " t =1 "
- Indice t 1= Gt / G1x100
- La période t =1 s'appelle la période de base ou période de référence
- Exemple :
- Le prix du kilogramme de viande au niveau de la wilaya d'Alger est de 1000 DA en 2005 et de 1100 DA en 2006.
- Indice 2005 / 2006 = prix 2006/ prix 2005 = 1100 / 1000 =1,1 ou 110.
- cela signifie que le prix de la viande est égal à 1,1 fois celui de l'année 2005.
- Si on utilise le taux de croissance, cela signifie que le prix de la viande a augmenté de 10 % de 2005 à 2006.

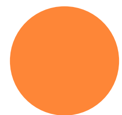
Effectif, fréquences relatives, fréquences relatives cumulées

- **Effectif**
- L'effectif total est le nombre d'individus dans la série.
- L'effectif ou fréquence absolue est le nombre d'individus appartenant à une modalité donnée.
- Prenons l'exemple de la distribution de 50 malades selon le sexe. Parmi ces 50 malades, 15 sont de sexe masculin et 35 de sexe féminin. Les effectifs correspondant à chacune des deux modalités sont 15 et 35.

Fréquence relative

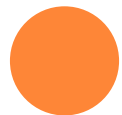
- La fréquence relative est le rapport entre l'effectif de la classe ou de la modalité et l'effectif total de la série, le numérateur fait obligatoirement partie du dénominateur.
- Elle est exprimée le plus souvent en %.

Exemple :



EXEMPLE (TABLEAU)

Fréquence cardiaque battement / minute	Effectif (ni)	Fréquence relative	Fréquence relative/100	Fréquence relative cumulée
50 – 59	2	0,2	20 %	20 %
60 – 69	3	0,3	30 %	50 %
70 – 79	4	0,4	40 %	90 %
80 – 89	1	0,1	10 %	100 %
N	10	1,0	100	



EFFECTIF, FRÉQUENCES RELATIVES, FRÉQUENCES RELATIVES CUMULÉES

- Fréquence Relative

  Effectif de la classe
- FR = ------------------------------------

  Effectif total

- Classe = 60 – 69 battement /minute : FR = 3/10 = 0,3 x 100 = 30 %
- 30 % des médecins ont une fréquence cardiaque comprise entre 60 et 69 battements cardiaques par minute
- **L'effectif total** est le nombre d'individus dans la série.
- L'effectif ou fréquence absolue est le nombre d'individus appartenant à une modalité donnée.
- Prenons l'exemple de la distribution de 50 malades selon le sexe. Parmi ces 50 malades, 15 sont de sexe masculin et 35 de sexe féminin. Les effectifs correspondant à chacune des deux modalités sont 15 et 35.
- **L'effectif cumulé** :
- Les fréquences cumulées peuvent donner de manière instantanée des informations très utiles. Elles se calculent aussi bien pour les effectifs que pour les fréquences relatives.

EFFECTIF, FRÉQUENCES RELATIVES, FRÉQUENCES RELATIVES CUMULÉES

- Lorsque les classes d'une variable sont ordonnées, on peut obtenir les effectifs cumulés. L'effectif cumulé croissant - plus de - (ou décroissant – moins de -) d'une valeur x_i ou d'une classe est la somme des effectifs correspondant à cette valeur ou à cette classe et des valeurs ou classes inférieures (ou supérieures).

- Pour calculer les effectifs cumulés croissants, on compte à partir de la borne supérieure, on additionne au cumul les chiffres des lignes précédentes.
- Pour les effectifs cumulés décroissants, on compte à partir de la borne inférieure, on enlève à l'effectif total les chiffres des lignes précédentes.
- Le principe est le même pour les fréquences cumulées. Les fréquences cumulées peuvent aussi être obtenues en divisant les effectifs cumulés par l'effectif total de la série.
- Ces effectifs cumulés permettent de dire combien d'individus ont une valeur supérieure ou inférieure à une valeur donnée.

Exemple : Soit la distribution statistique suivante : Âge des étudiants

Age (ans)	xi	ni	Effectif Cumulé	Fréquence Relative %	Fréquence Relative cumulée %
22 – 24	23	1	1	20	20
24 – 26	25	3	4	60	80
26 – 28	27	1	5	20	100
Total		5		100	Total

II.-SOURCES D'INFORMATIONS ET BIAIS

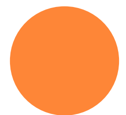
Sources d'informations:

Le recensement:

Le recensement consiste à recueillir les données démographiques, économiques, et sociales se rapportant à un moment donné à tous les habitants d'un pays donné.

Le sondage :

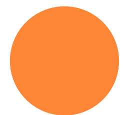
- Le sondage consiste à déterminer certains caractères d'une population à partir des résultats obtenus sur une partie de la population, appelée **échantillon**
- La définition d'une procédure d'échantillonnage repose sur les principes suivants
- La définition de **l'unité de sondage**
- **L**a **taille** de l'échantillon
- Le choix des **individus**



II.-Sources d'informations et Biais

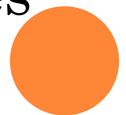
- **Les organisations professionnelles liées aux entreprises**
- **Les organismes nationaux:**

- MSPRH
- INSP
- IP
- l' Agence Nationale de documentation en santé (ANDS),
- le Centre national de toxicologie,
- l'Agence Nationale du Sang,
- l' Ecole Nationale de Santé Publique,
- l'Institut national de pédagogie,
- le Laboratoire National de contrôle des produits pharmaceutiques,
- le Centre national de pharmacovigilance (Formation à la prescription rationnelle des médicaments)
- la caisse nationale de sécurité sociale,
- Office National des Statistiques



II.-Sources d'informations et Biais

- **Les organismes internationaux :**
- ONU,
- OMS,
- l'Organisation des Nations Unis pour l'Education, la Science et la Culture (UNESCO),
- le Bureau International du Travail (BIT),
- la Communauté Economique Européenne(CEE)
- **Les organismes para officiels et privés**
- **Les enquêtes et questionnaires**
- Lorsque les sources d'information sont insuffisantes, on peut envisager d'entreprendre des études particulières compte tenu des objectifs retenus.



II.-Sources d'informations et Biais

Biais

- Pour tirer des conclusions valables d'étude statistique, il faut identifier les biais. Ils peuvent être liés à des
- **Erreurs d'échantillonnage** (intervalle de confiance, degré de précision recherché)
- Plus la précision est élevée, plus la taille de l'échantillon doit être étendue.
- Le degré de confiance accepté généralement est de 95% c'est-à-dire qu'on accepte qu'il y ait 5 chances sur 100 pour que la vérité s'écarte de plus de x % du résultat trouvé.
- **Erreurs dans le recueil des données**
- Erreurs dues aux questionnaires (questions mal rédigées)
- Erreurs dues aux enquêtés (oubli, crainte d'avouer son ignorance, ne pas communiquer toute l'information dont il dispose
- Erreurs dues aux enquêteurs (mauvaise maitrise du questionnaire)

Types de VARIABLES

QUALITATIVES

Variables Non Mesurables

Sexe,
groupe sanguin,
nationalité,...

QUANTITATIVES

Variables pouvant être QUANTIFIEES (mesurables)

discrètes ou (discontinues)

Parité,
nombre d 'enfants
nombre de lits,...

Continues

Hb, glycémie, taille,
poids, ,...

Echelle de classification de ces variables

Echelle NOMINALE

Les classes sont NOMMEES sans ordre Déterminé

Echelle ORDINALE

Les classes sont NOMMEES DANS UN ORDRE DETERMINE

Echelle NUMERIQUE

-V.Numériques

-Mesurées

Echelle NOMINALE



Les classes sont NOMMEES sans ordre déterminé

Exemple :

Groupes sanguins

profession

État civil

La plus simple: DICHOTOMIQUE :

malade / non malade

masc / fém

oui / non

Echelle ORDINALE



Les classes sont NOMMEES DANS UN ORDRE DETERMINE

Exemples :

- Niveau d 'instruction
- revenus
- Catégorie professionnelle

Echelle NUMERIQUE

Discontinue ou discrète

CONTINUE

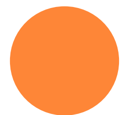
Les observations sont des nombres ENTIERS

Chaque observation peut prendre n'importe quelle valeur numérique (infini)

Exemples: freq.card., parité, nbre de vaccin, âge en année révolue,...

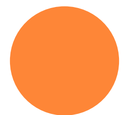
Exemples: glycémie, poids, taille, ...

- Les variables quantitatives peuvent être transformées en variables ordinales puis en variables qualitatives ; une variable qualitative ne peut pas être transformée en variable quantitative : poids (kilogramme - ou poids léger, moyen, lourd..).



IV/MISE EN ORDRE DES DONNÉES ET REGROUPEMENT DES VALEURS EN CLASSES

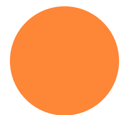
- Lors du recueil des données, les observations se trouvent sans ordre, il est nécessaire, afin de pouvoir les décrire, de les trier, de les regrouper, de les classer.
- On procède d'abord au tri et à la mise en ordre des données.
- La mise en ordre des données est l'étape qui suit immédiatement celle du recueil des valeurs de la (ou des) variable(s) étudiée(s).



IV/MISE EN ORDRE DES DONNÉES ET REGROUPEMENT DES VALEURS EN CLASSES

- **Pour les variables qualitatives**, il y a autant de catégories qu'il y a de modalités de la variable.
- Tableau 1.1: Distribution de 50 malades selon le sexe

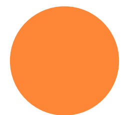
Sexe	Effectif	%
Masculin	15	30
Féminin	35	70
Total	50	100



IV/Mise en ordre des données et regroupement des valeurs en classes

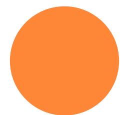
Nombre d'épisodes	Effectif	%
0	3	15.8
1	7	36.8
2	6	31.6
3	2	10.5
4	1	5.3
Total	19	100

- **Pour les variables quantitatives discrètes**, il y a autant de catégories que de valeur précise de la variable si les modalités sont peu nombreuses.
- On a observé parmi 19 personnes, le nombre d'épisodes de syndrome grippal pendant un an. Le nombre d'épisodes variant de 0 à 4, on classe les 19 individus selon les valeurs entières successives 0, 1, 2, 3, 4
- On peut aussi réunir les deux dernières modalités sous une rubrique unique « 3 et plus ». L'effectif correspondant à cette modalité est 2 + 1 = 3.



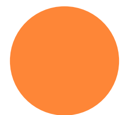
IV/MISE EN ORDRE DES DONNÉES ET REGROUPEMENT DES VALEURS EN CLASSES

- Pour présenter les données d'une **variable quantitative** sur un nombre important d'individus, on procède à la transformation d'une variable quantitative continue en une variable quantitative discrète (classes) ou d'une variable quantitative discrète en une variable qualitative ordinale (ordre) ; cette opération fait perdre de l'information.
- Pour pouvoir dresser un tableau d'effectifs qui soit facile à lire, il faut grouper les observations ou données dans un certain nombre de classes successives, contigües ne se recouvrant pas.



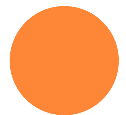
IV/MISE EN ORDRE DES DONNÉES ET REGROUPEMENT DES VALEURS EN CLASSES

- Chaque classe est définie par ses limites, son amplitude et sa valeur centrale.
- Les limites de classes doivent être bien précisées. On définit des bornes (limites) de classe, la borne inférieure est la plus petite valeur, la borne supérieure est la plus grande, chaque classe comprend toutes les valeurs égales ou supérieures à sa borne inférieure mais uniquement les valeurs inférieures à sa limite supérieure.
- Une observation ne doit être située que dans une seule classe. C'est pour cela que la convention adoptée est de toujours inclure dans la classe la limite inférieure et donc de toujours exclure la limite supérieure.
- Les classes sont contiguës, sans chevauchement, borne inférieure comprise, borne supérieure exclue.
- L'amplitude de classe est la différence entre la borne supérieure et inférieure de la classe,
- Le nombre de classes est en général déterminé de façon arbitraire ou en utilisant des règles. Il n'existe pas de règle permettant d'imposer le nombre de classes. Souvent, on s'efforce de construire des classes d'amplitude égale. Si on choisit une amplitude égale pour les classes, la règle suivante a pu être proposée : Nombre de classes = étendue / amplitude.
- L'étendue (ou marge) est la différence entre la valeur la plus grande et la valeur la plus faible de la série.



Comment déterminer le nombre de classes

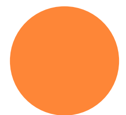
- Diverses formules empiriques permettent d'établir le **nombre de classes pour un échantillon** de taille *n*.
- La règle de STURGE : **Nombre de classes = 1+3logN**
- La règle de YULE : **Nombre de classes =** $2,5\ N^{\frac{1}{4}}$
- **En général on utilise K= racine carrée de N**
- **Nombre de classes = Etendue/amplitude (même amplitude)**



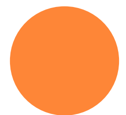
Une classe est définie par:

- limites de classes :

  x_{max} et x_{min}
- centre de classe
- amplitude de classe (intervalle de classe ou étendue de classe):
  - de préférence EGALE mais pas obligatoire



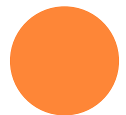
- Exemple : On a regroupé les données par classes des fréquences cardiaques de 10 médecins
- 59, 72, 58, 65, 77, 83, 72, 77, 62,62
- **Estimation du nombre de classes = 1 + 3 log N**
- Si N = 10, nombre de classes = 1 + 3 (1) = 4 classes
- si N = 100, nombre de classes = 1 + 3 (2) = 7classes



- Le groupement des données en classes revient à assimiler toutes les valeurs de la classe à une seule valeur, le centre de classe. La valeur centrale de la classe est le centre de la classe (en règle générale, le centre de classe est un nombre entier ou un nombre avec peu de décimales). Le centre de classe est la valeur située au milieu de la classe (limite inférieure de la classe + limite supérieure classe) / 2.
- Par exemple, on a relevé le poids de 19 étudiants. L'unité de poids retenue est le kilogramme et les résultats sont les suivants :
- 76,340 60,400 68,280 57,740 64,990 83,450 79,650
  64,100 72,880 69,120 59,990 61,820 61,820 76,360
  66,330 52,990 70,560 70,130 65,450
- L'étendue dans l'exemple du poids est : 83.450 – 52.990 = 30.460 kg
- Si on décide de prendre une amplitude de 5kg, le nombre de classes est environ de : 30.46 / 5 = 6,09 classes.
- Souvent, on s'efforce de construire des classes d'amplitude égale

Le tableau d'effectifs peut être présenté comme suit lorsqu'on retient 7 classes.

Poids (Kg)	Centre de classes	Effectif	%
50-54	52.5	1	5.3
55-59	57.5	2	10.5
60-64	62.5	5	26.3
65-69	67.5	4	21.2
70-74	72.5	3	15.8
75-79	77.5	3	15.8
80-84	82.5	1	5.3
Total		19	100.0



VIII/ REPRÉSENTATION TABULAIRE

- Un tableau doit être simple, il doit comporter toutes les informations nécessaires à sa compréhension, il se suffit à lui-même.
- Le tableau comporte un en-tête, des têtes de colonnes, des têtes de lignes, un corps.
- Le total des effectifs de la série et le total des fréquences sont mentionnés.
- Titres et sous titres du tableau indiquent le contenu du tableau et doivent se trouver au dessus du tableau.
- Le titre est clair et concis, il précise les caractéristiques de temps, de lieu et de personnes (quoi, où, quand, comment ?). Il est précédé du numéro du tableau.
- Les titres des lignes et des colonnes sont clairement libellés. Ils précisent le contenu du tableau statistique
- Les unités de mesure doivent être mentionnées en tête de la ligne ou de la colonne correspondante ;
- les codes, abréviations, symboles sont expliqués en bas de page de même que la source pour des données originales.

Tableau statistique à simple entrée

Nombre d'enfants à charge (xi)	Effectifs (ni)	%
0	5	
1	17	
2	31	
3	20	
4	11	
5	4	
6	1	
Total :N	89	

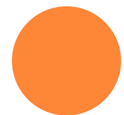
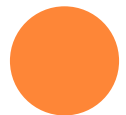


Tableau statistique à double entrée

Tranches d'âge (ans) / Classes poids (kg)	3-4	4-5	5-6	Total
< 15	19	7	1	27
15-20	32	21	12	65
20-25	3	18	28	49
25-30			1	1
Total	54	46	42	142

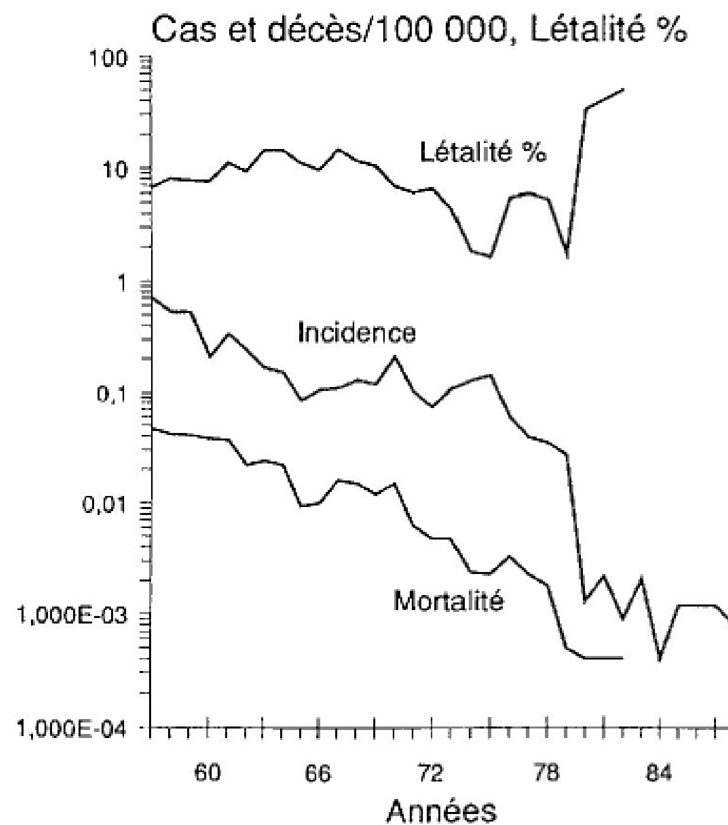


Méthodes graphiques

visualisation de la distribution

comparaison avec d 'autres distributions

choix des paramètres de réduction



Source : Centers for Disease Control, Atlanta, USA.

Généralités sur la représentation graphique (3)

Variables quantitatives continues :

surtout HISTOGRAMME et POLYGONES de FREQUENCE

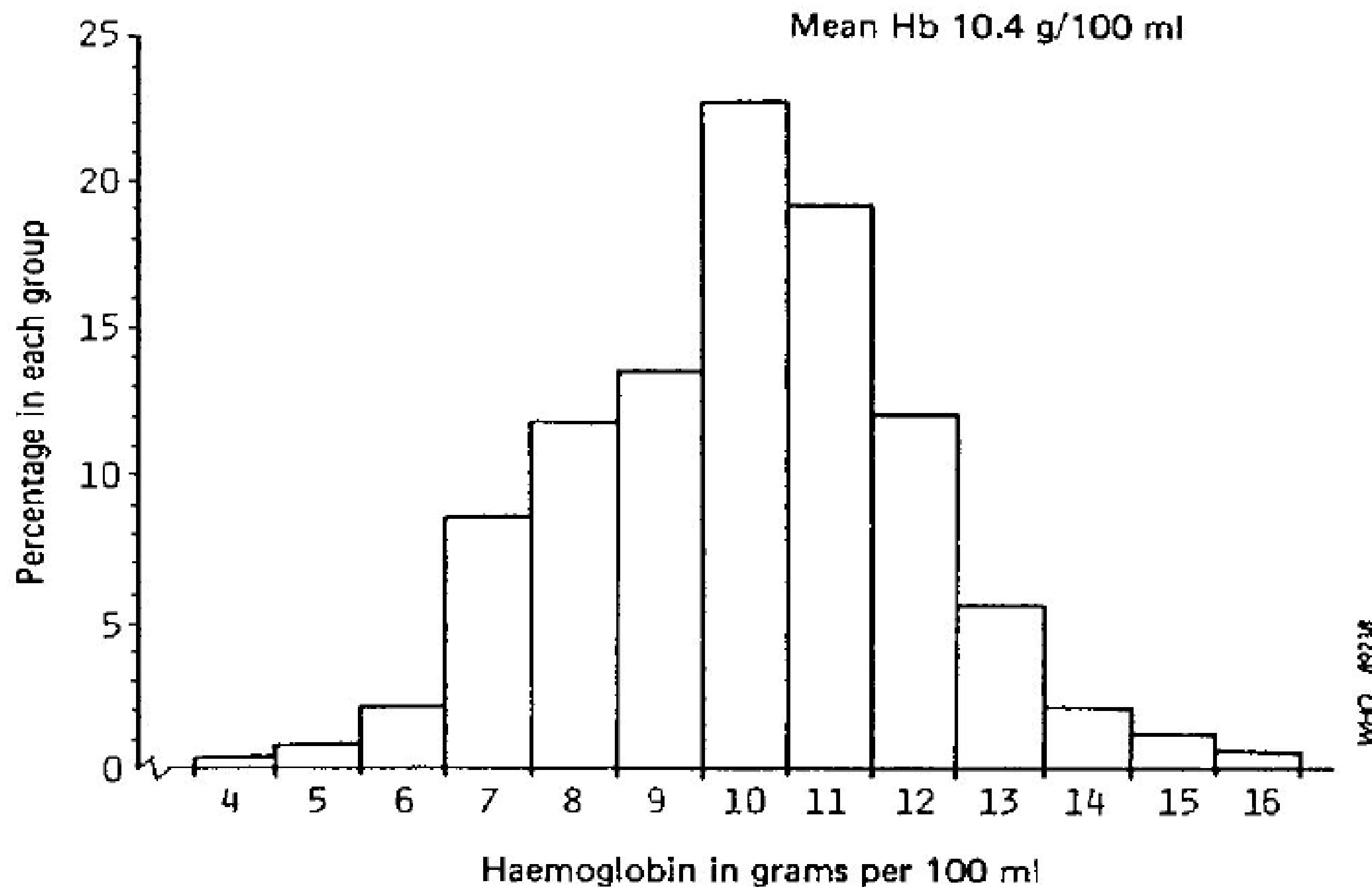
variables quantitatives comptées et qualitatives :

de préférence DIAGRAMMES

HISTOGRAMME

- Histogramme présente une distribution de fréquence avec des barres verticales CONTIGÜES (ne pas confondre avec **diagramme en bâton** où il y a un **espace** entre les bâton)
- la SURFACE de chaque barre est proportionnelle à la valeur qu'elle représente
- SURFACE TOTALE = 100 % = 1
- quelques termes :
  - **EFFECTIF** (= fréquence absolue) = nombre d 'observation dans une classe
  - **FREQUENCE** (relative) =
    - effectif classe i / effectif Total
  - **FREQUENCE CUMULEE** = somme de la fréquence d 'une classe et de toutes celles qui la précède.

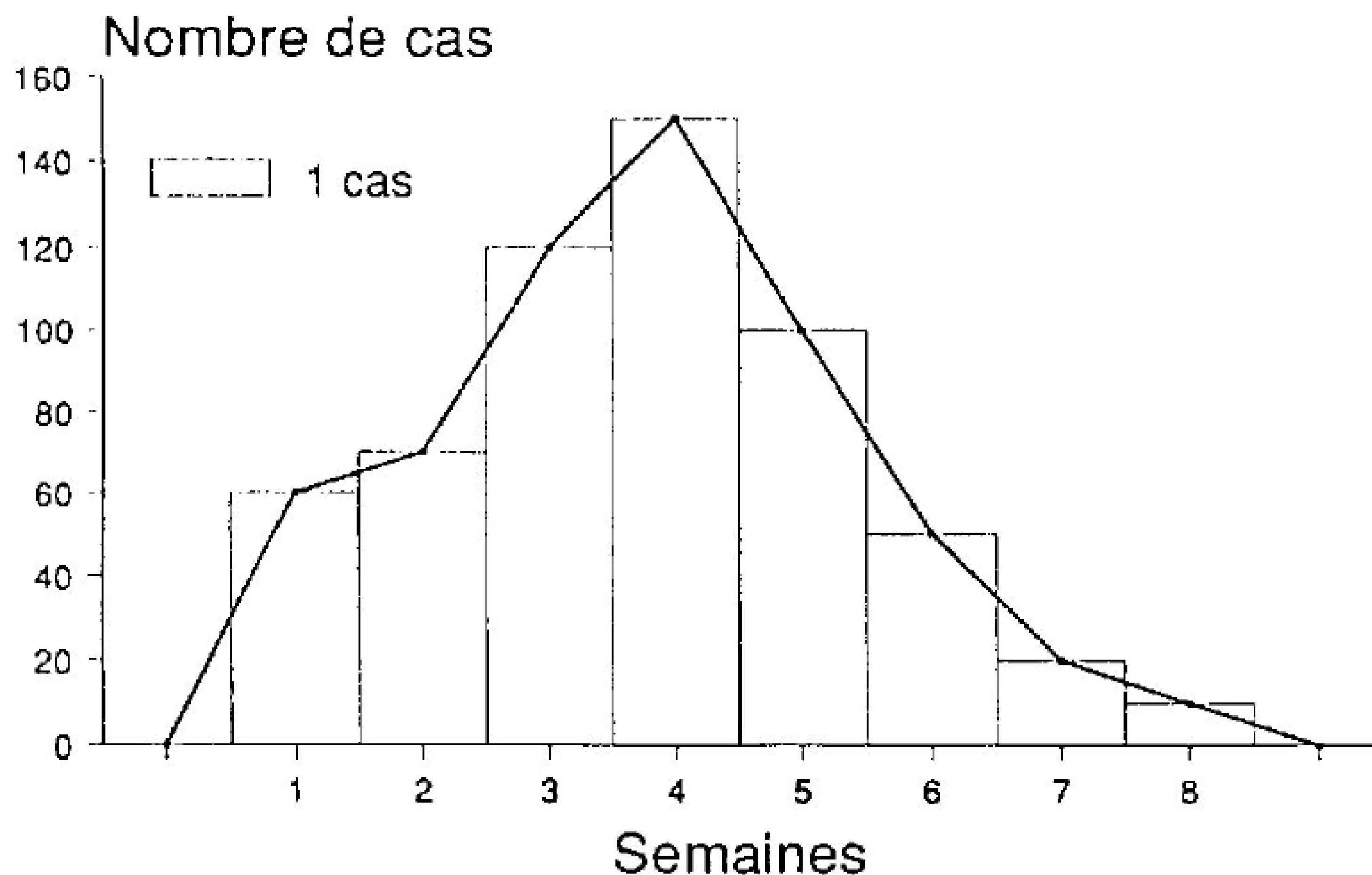
HISTOGRAMME



POLYGONE DE FREQUENCE

- Construction:
  - **point de départ = histogramme**
  - **On joint simplement par une ligne le milieu des bords supérieurs des rectangles représentant chaque classe d 'un histogramme**
  - **la surface sous le polygone = surface de l 'histogramme**
  - **ATTENTION à la FERMETURE du polygone**
- c 'est la méthode indiquée pour représenter plusieurs séries de données sous la forme d 'une distribution de fréquence

Figure 5 b : Nombre de cas de syndromes grippaux par semaine, Echantillonville.

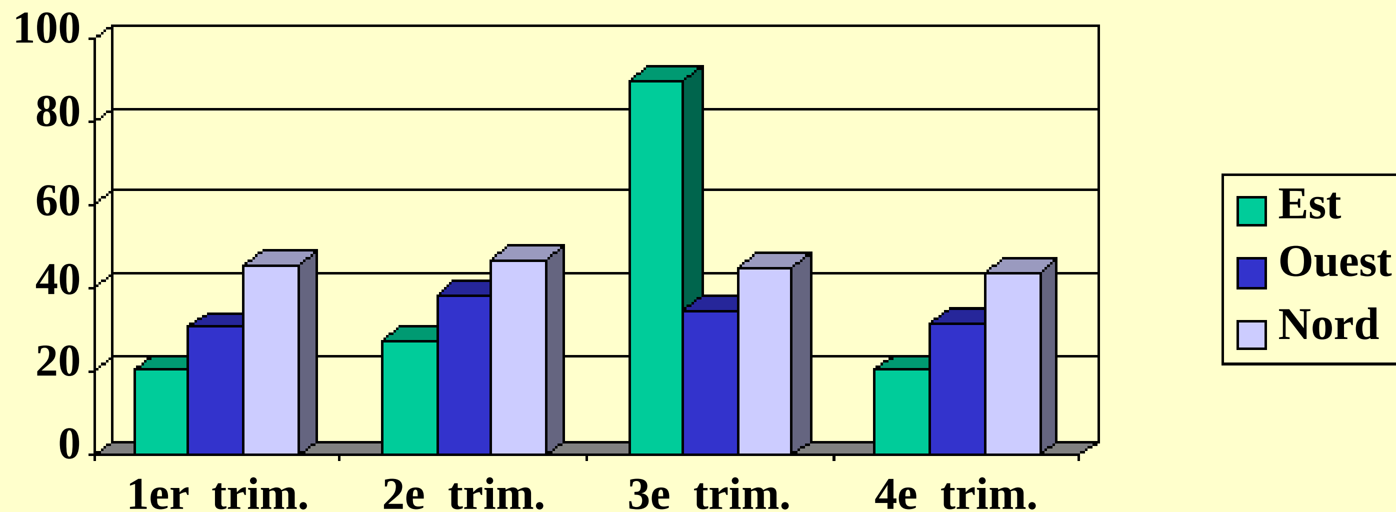


DIAGRAMMES:

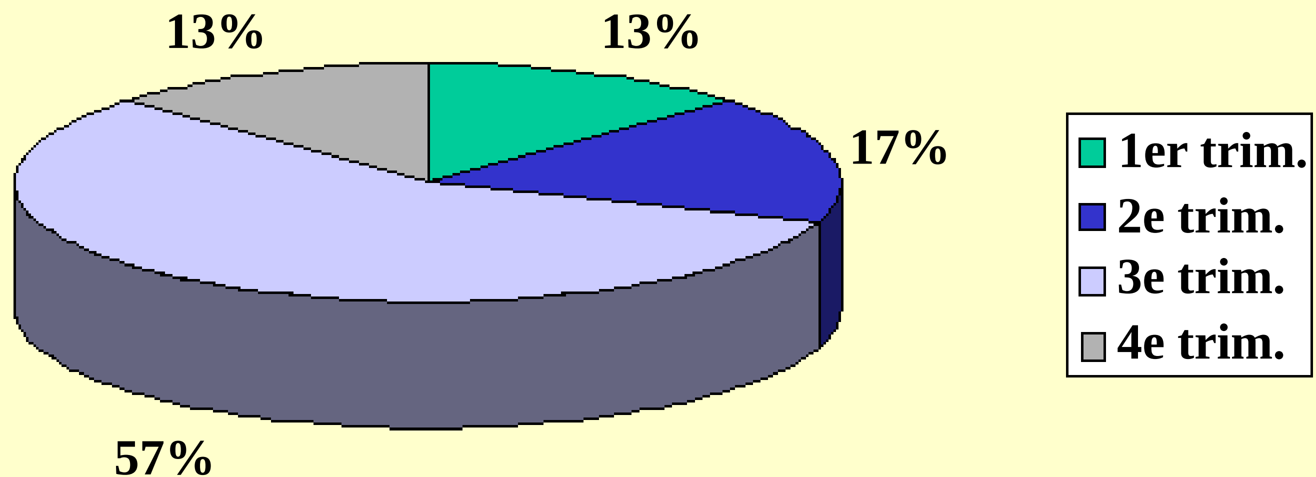
pour les variables quantitatives comptées et les variables qualitatives

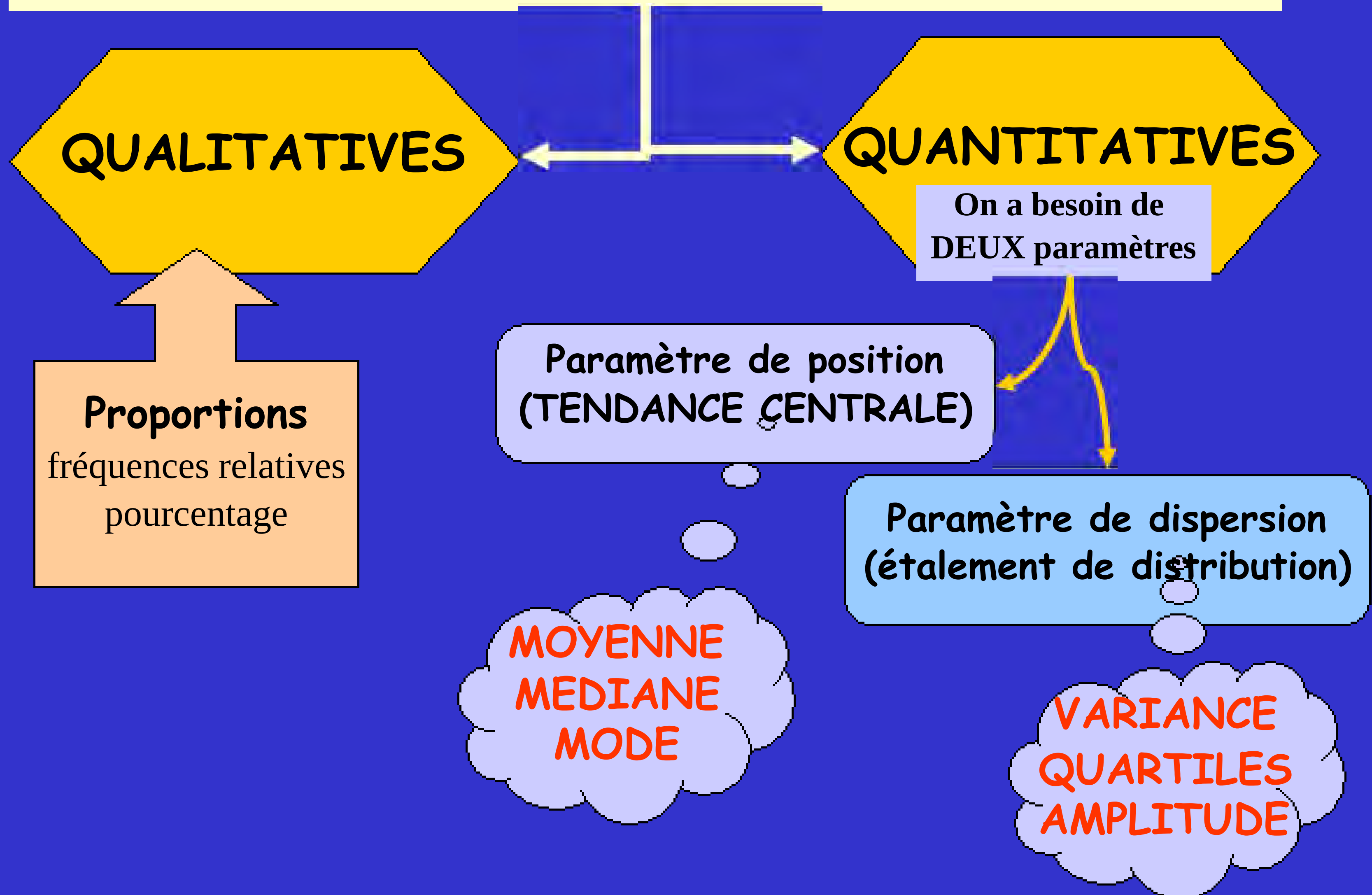
- Le plus répandu : diagramme en bâtons (ou en barres)
- pie (camembert, tarte, …) : quand on fait référence au TOTAL de 100%

Diagramme en bâton



Pie / camembert / tarte




Résumé des données
QUALITATIVES
QUANTITATIVES
On a besoin de DEUX paramètres
Proportions
fréquences relatives
pourcentage
Paramètre de position (TENDANCE CENTRALE)
Paramètre de dispersion (étalement de distribution)
MOYENNE
MEDIANE
MODE
VARIANCE
QUARTILES
AMPLITUDE

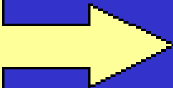
Paramètres de tendance centrale

- la MOYENNE (arithmétique)
- la MEDIANE
- le MODE

Paramètres de DISPERSION

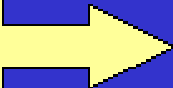
À chaque paramètre de POSITION…………un paramètre de DISPERSION

MOYENNE



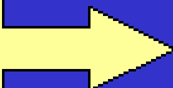
VARIANCE et ECART TYPE

MEDIANE



QUARTILES

MODE



AMPLITUDE

V/ CARACTÉRISTIQUES DE POSITION

- **Quantiles**
  Les quartiles divisent la série statistique en quatre parties égales comprenant le même nombre de sujets.
- Le premier quartile (quartile inférieur) est la valeur de la variable du $25^{\text{ème}}$ sujet sur 100.
- Le deuxième quartile n'est autre que la médiane, c'est la valeur de la variable du $50^{\text{ème}}$ sujet sur 100.
- Le troisième quartile (quartile supérieur) est la valeur de la variable du $75^{\text{ème}}$ sujet sur 100.

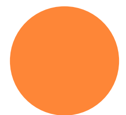
- Dans une série de 60 cas, par exemple, le rang du premier quartile sera 15, le rang du deuxième quartile sera 30, le rang du troisième quartile sera 45. Une fois le rang déterminé, on recherche les valeurs correspondantes de la variable.

- Pour le poids de 19 étudiants, le rang du premier quartile est 4,8, le rang du deuxième quartile est 9,5, le rang du troisième quartile est 14,3.
- On peut à priori affecter aux différents quartiles les valeurs des centres de classes dans lesquelles ils se trouvent, soit pour notre exemple et par ordre : 62,5, 67,5, 72,5 Kg.
- Il est intéressant de représenter sur graphique la distribution des fréquences relatives cumulées d'une série d'observation ceci nous permet de déterminer graphiquement la médiane.

V/ CARACTÉRISTIQUES DE POSITION

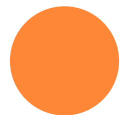
- **Quartile 1** = n + 1 /4 --------- 25 % des observations < inférieures
  75% des observations > supérieures
- **Quartile 2** = médiane
- **Quartile 3** = (n + 1) x 3 / 4 75 % des observations < inférieures
  25% des observations > supérieures

- **Déciles :**
- Les déciles sont au nombre de 9. Ce sont des valeurs de la variable qui partagent la série statistique en 10 parties comprenant chacune 1/10ème de l'effectif total.
- Le premier décile est la note du 10ème sujet sur 100.
- Le deuxième décile est la note du 20ème sujet sur 100
- Le cinquième décile se confond avec le deuxième quartile et la médiane
- Le décilage est d'un usage fréquent, notamment en biométrie, parce qu'il permet de situer rapidement et facilement la position d'un sujet quelconque par rapport aux autres sujets de la série.



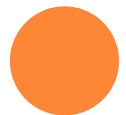
V/ CARACTÉRISTIQUES DE POSITION

- **Percentiles :**
- Les percentiles sont au nombre de 99. Ce sont des valeurs de la variable qui partagent la série statistique en 100 parties comprenant chacune 1/100ème de l'effectif total.
- De façon générale, les percentiles sont utilisés lorsque le nombre de valeurs de la série statistique est supérieur à 1000.
- Le 42ème percentile, par exemple, est la note du 420ème sujet sur 1000.



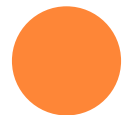
VI /LES PARAMÈTRES DE TENDANCE CENTRALE

- **VI .1. Moyenne**
- C'est la mesure de tendance centrale la plus utilisée, elle résume une série de données d'une variable quantitative.
- La moyenne est la somme algébrique des valeurs mesurées divisée par le
- nombre de valeurs mesurées, d'individus, d'observations.
- $m = \sum x / N$
- La moyenne est facile à calculer. Son inconvénient est d'être sensible et influencée par les valeurs extrêmes.
- Exemple :
- Caractère étudié : fréquence cardiaque des 10 médecins
- Résultats : 59, 72, 58, 65, 77, 83, 72, 77, 62, 62



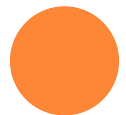
MOYENNE : OBSERVATIONS NON GROUPÉES

Valeur du caractére x	Effectif
58	1
59	1
62	1
62	1
65	1
72	1
72	1
77	1
77	1
83	1
Total = 687	10



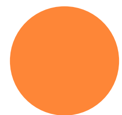
Moyenne : Observations non groupées

- $m = \Sigma xi / N = 687 / 10 = X^{-} = 68{,}7$ battements /mn
- La fréquence cardiaque moyenne des 10 médecins est de 68,7 battements /mn



MOYENNE: OBSERVATIONS GROUPÉES

Fréquence cardiaque	Valeur centrale (Xi)	Effectif (Ni)	NiXi
55-59	57.5	2	115.0
60-64	62.5	2	125.0
65-69	67.5	1	67.5
70-74	72.5	2	145.0
75-79	77.5	2	155.0
80-84	82.5	1	82.5
Total		10	690.0



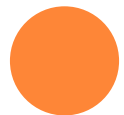
MOYENNE: OBSERVATIONS GROUPÉES

- m = Σni xi / N = 690 / 10 = 69 battements / minute
- La fréquence cardiaque moyenne des 10 médecins est de 69 battements /mn

- Les caractéristiques de tendance centrale et de position sont insuffisantes pour caractériser complètement une série statistique, c'est-à-dire décrire sa structure interne ou le mode de dispersion des valeurs de la série statistique autour d'une caractéristique centrale.
- Deux distributions peuvent avoir la même valeur de tendance centrale sans être réparties de la même manière.
- Ex :

3 .4.6.14.18.19.20. moyenne =12, médiane =14

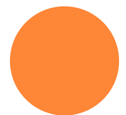
5.7.10.14.15.16.17 moyenne =12, médiane =14



MODE

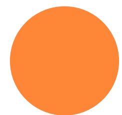
- **Mode :** est la valeur de la série la plus fréquente, c'est la valeur qui est observée le plus fréquemment

- **Mode** : effectif le plus élevé



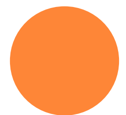
MODE: VARIABLES QUANTITATIVES DISCONTINUES

Nombre d'épisodes	Effectif
0	3
1	**7**
2	6
3	2
4	1



MODE: VARIABLES QUANTITATIVES CONTINUES (POIDS DE 19 ÉTUDIANTS)

Poids (Kg)	Centre de classes	Effectif	%
50-54	52.5	1	5.3
55-59	57.5	2	10.5
60-64	62.5	**5**	26.3
65-69	67.5	4	21.2
70-74	72.5	3	15.8
75-79	77.5	3	15.8
80-84	82.5	1	5.3
Total		19	100.0



VI .3. MÉDIANE

- La médiane est la valeur qui sépare une série statistique en deux parties égales telle qu'une moitié des observations rangées lui est inférieure et l'autre moitié lui est supérieure.
- Médiane = L_m (n/2 - f cum)

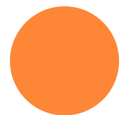
L_m : limite inférieure de la classe dans laquelle se trouve le n/2 élément

f_m: fréquence de l'élément médian

i : amplitude de classe

n : effectif de l'échantillon

f cum : fréquence cumulée jusqu'à la limite inférieure de la classe médiane



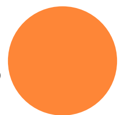
VI .3. MÉDIANE

- **Calcul de la médiane pour les variables quantitatives discontinues**
- On détermine d'abord la position de la médiane.
- Si N est impair :
- Position de la médiane : = N+1/ 2
- Si N est pair : Il faut prendre la moyenne entre les deux observations
- N N
- ---- et ----- +1
- 2 2

Exemple

- N = 10
- N représente 10 médecins chez lesquels on a mesuré la fréquence cardiaque dont voici les résultats :
- 58, 59, 62, 62, 65, 72, 72, 77, 77,83.
- On déterminera la position de la médiane qui est placée entre les observations
- N 10 N
- --- = ----= 5 et ---- + 1= 6 65+72
- 2 2 2 -------- = 68.5 battements par minute

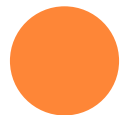
  2
- Commentaire : 50% des médecins présente une fréquence cardiaque inférieure à 68.5 battements par minute.



VI .3. MÉDIANE

- **<u>Calcul de la médiane pour les séries classées</u>**
- la série statistique suivante représente 35 élèves chez lesquels on a mesuré la taille en m .

Classes: taille (m)	Effectif
1.55-1.60	1
1.60-1.64	6
1.65-1.69	8
1.70-1.74	10
1.75-1.79	7
1.80-1.84	1
1.85-1.89	2
Total (N)	35



VI .3. MÉDIANE

- effectifs1C : effectifs cumulés des classes inférieures à la classe médiane :=15
- Effectif de la classe médiane= effectif de la classe médiane=10
- C= valeur de l'amplitude de la classe médiane
- L1= Borne inférieure de la classe médiane

Valeur de la médiane = L1+ { [N /2 – eff1C / Eff médian] X C }

- Valeur de la médiane= **1.70 + [17,5-15 /10) × 0,04]= 1,71 m**
- 50 % des élèves ont une taille inférieure à 1,71 cm.

VII. Paramètres de dispersion

- **VII.1. L'écart-type et la variance**
- L'écart-type et la variance sont en réalité deux expressions du même paramètre de dispersion. La variance est simplement le carré de l'écart-type.
- La variance est la moyenne des carrés des écarts à la moyenne. Elle est généralement notée S^2 lorsqu'elle est calculée à partir d'un échantillon.
- En cas de données non groupées (individuelles), la variance s'écrit :

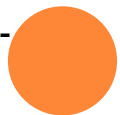
$$S^2 = [\sum (x-m)^2] / (N)$$

- En cas de données groupées, et de manière générale, la variance s'écrit :

$$S^2 = [\sum n(x-m)^2] / (N)$$

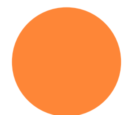
$$S^2 = \left(\frac{\sum n_i x_i^2}{N}\right) - (m^2)$$

- **L'écart type** est la caractéristique de dispersion la plus utilisée car la plus satisfaisante.
- La variance et l'écart-type renseignent tous deux sur l'étalement de la série statistique autour de la moyenne. Mais tandis que la variance mesure le carré d'une grandeur, l'écart-type d'une distribution s'exprime dans la même unité de mesure que les observations. Si par exemple, une variable s'exprime en Kg, la variance s'exprime en Kg^2 tandis que l'écart-type en Kg.



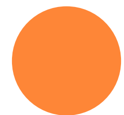
VII. Paramètres de dispersion

Taille (Xi)	ni	nixi	x-m	$(x-m)^2$	$(x-m)^2 x ni$
140-144	3				
144-148	17				
148-152	63				
152-156	82				
156-160	69				
160-164	31				
164-168	20				
168-172	4				
172-174	1				
174-178	1				
Total					



VII. Paramètres de dispersion

Taille (Xi)	ni	nixi	x-m	(x-m)²	(x-m)²×ni
140-144	3	426		179.56	538.68
144-148	17	2448		88,5481	1505,3177
148-152	63	9450		29.16	1837.08
152-156	82	12628		1.96	160.72
156-160	69	10902		6.76	466.44
160-164	31	5022		43.56	1350.36
164-168	20	3320		112.36	2247.2
168-172	4	680		213.16	852.64
172-174	1	173		309.76	309.76
174-178	1	176		424.36	424.36
Total	291	45225			9688,6371

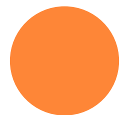


VII. Paramètres de dispersion

- X = Moyenne = 45259 / 291=155, 41 cm

$$\frac{\Sigma\, ni(x-m)^2}{N} = 9688{,}6371 / 291 = 33{,}29$$

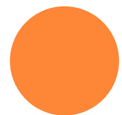
- Variance = 33,29
- T = √33,25 = 5,76
- La variabilité de la taille des adolescents par rapport à la taille moyenne est d'environ 6 cm.



VII.2.Le coefficient de variation

Observateurs	moyenne	E.Type	CV
A	75	7.5	10
B	75	3.75	5

- L'écart type mesure la variabilité des valeurs observées par rapport à la moyenne. On standardise ce paramètre de dispersion en le divisant par la moyenne.
- Coefficient variation = Ecart type/moyenne
- Exemple :
- 05 enfants sont mesurés par 02 observateurs A et B.
- Le travail de mensuration de l'observateur A est 2 fois moins précis que celui de l'observateur B



EXEMPLES

Exercice 1 :

Le taux de glucose sanguin (glycémie) déterminé chez 32 sujets est donné ci-dessous en g/l :

0.85 0.87 0.90 0.93 0.94 0.94 0.95 0.97 0.97 0.98 0.98 0.99
1.00 1.01 1.03 1.03 1.03 1.04 1.06 1.07 1.08 1.08 1.10 1.10
1.11 1.13 1.14 1.14 1.15 1.17 1.19 1.20

1) De quel type de variable s'agit-il ?
2) faire une répartition en classes, en justifiant le choix du nombre de classes et l'intervalle de classe.
3) Faire la représentation graphique.
4) Calculer la moyenne et l'écart type de cette série statistique.
5) Construire la courbe des effectifs cumulés croissants.
6) Déterminer la médiane et les quartiles de cette série par la méthode graphique.
7) Déterminer le mode ou la classe modale.



- **1) Variables quantitatives continues.**
- **2) N=32, la racine carrée de 32 nous donne environ 6 classes.**
- **Le tableau suivant va résumer les données**



Tableau résumant les données

Classes	Centre de classe	Effectifs	E. cum	Fréquence	F. cum
	xi	ni	ni cum	fi	fi cum
[0,85 ; 0,91[	0,88	3	3	0,09	0,09
[0,91 ; 0,97[	0,94	6	9	0,19	0,28
[0,97 ; 1,03[	1	8	17	0,25	0,53
[1,03 ; 1,09[	1,06	5	22	0,16	0,69
[1,09 ; 1,15[	1,12	6	28	0,19	0,88
[1,15 ; 1,21[	1,18	4	32	0,13	1,00
TOTAL		32		100%	

Moyenne	1,03
Ecart type	0,09



Histogramme

Distribution de la glycémie chez 32 sujets
(date et lieu)

fréquence (n)

9
8
7
6
5
4
3
2
1
0

3 6 8 5 6 4

[0,85 ; 0,91[[0,91 ; 0,97[[0,97 ; 1,03[[1,03 ; 1,09[[1,09 ; 1,15[[1,15 ; 1,21[

glycémie g/l : variable quantitative continue

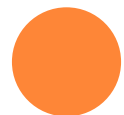


Diagramme cumulatif

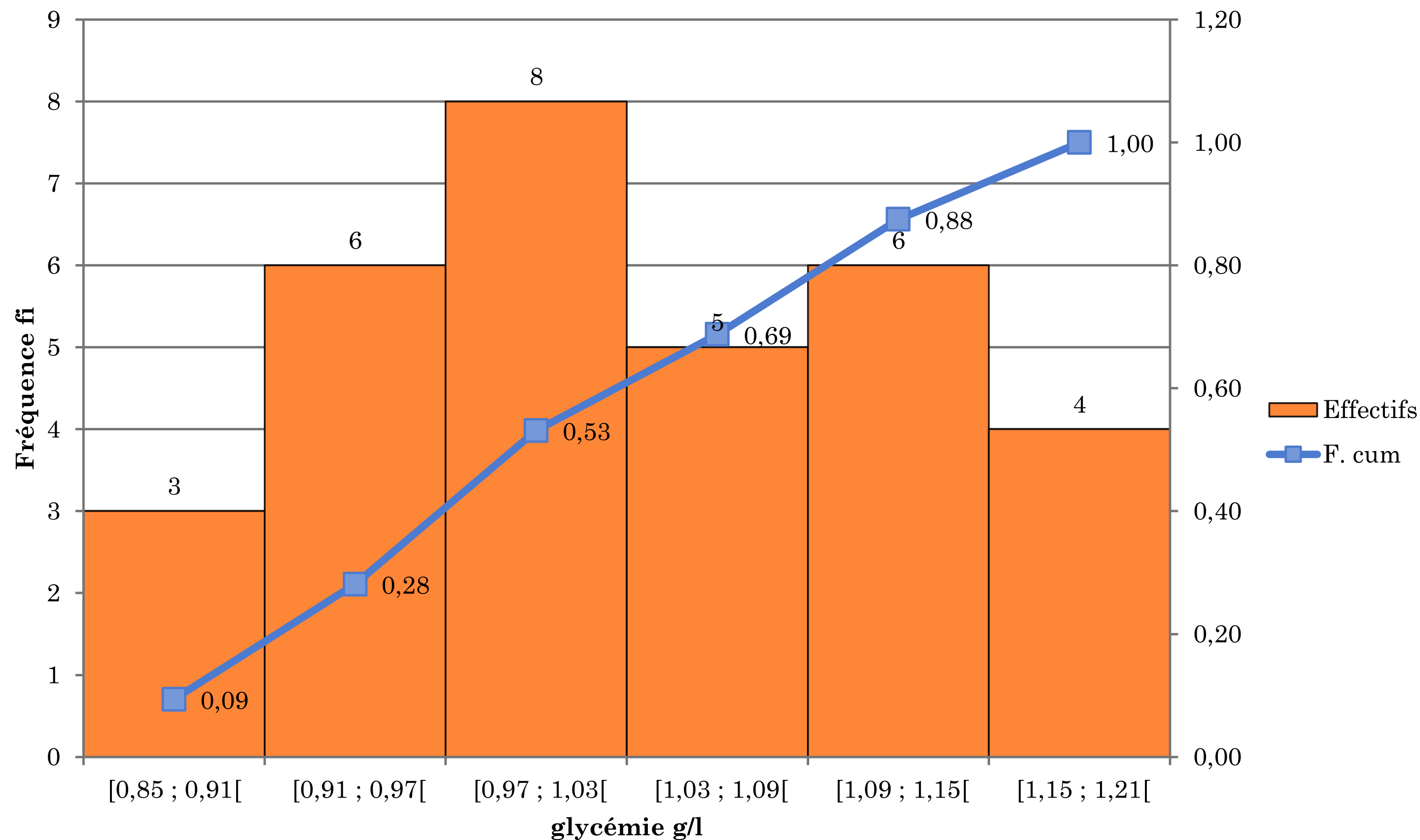


Diagramme circulaire: camembert

Affection	pourcçentage
Suspicion clinique	4%
Sida	6%
nouveaux nés	20%
femmes enceintes	70%

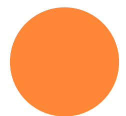
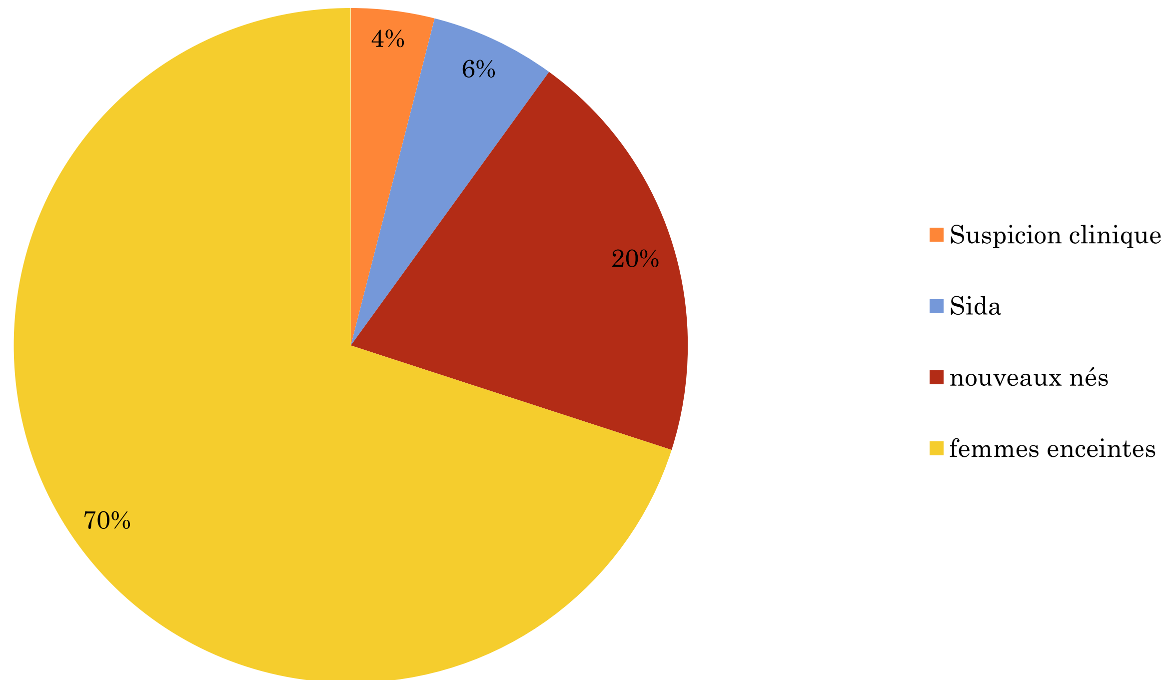


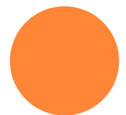
DIAGRAMME CIRCULAIRE: CAMEMBERT

Camembert: Distribution des érodiagnostics de toxoplasmose dans un laboratoire en fonction du type de patient

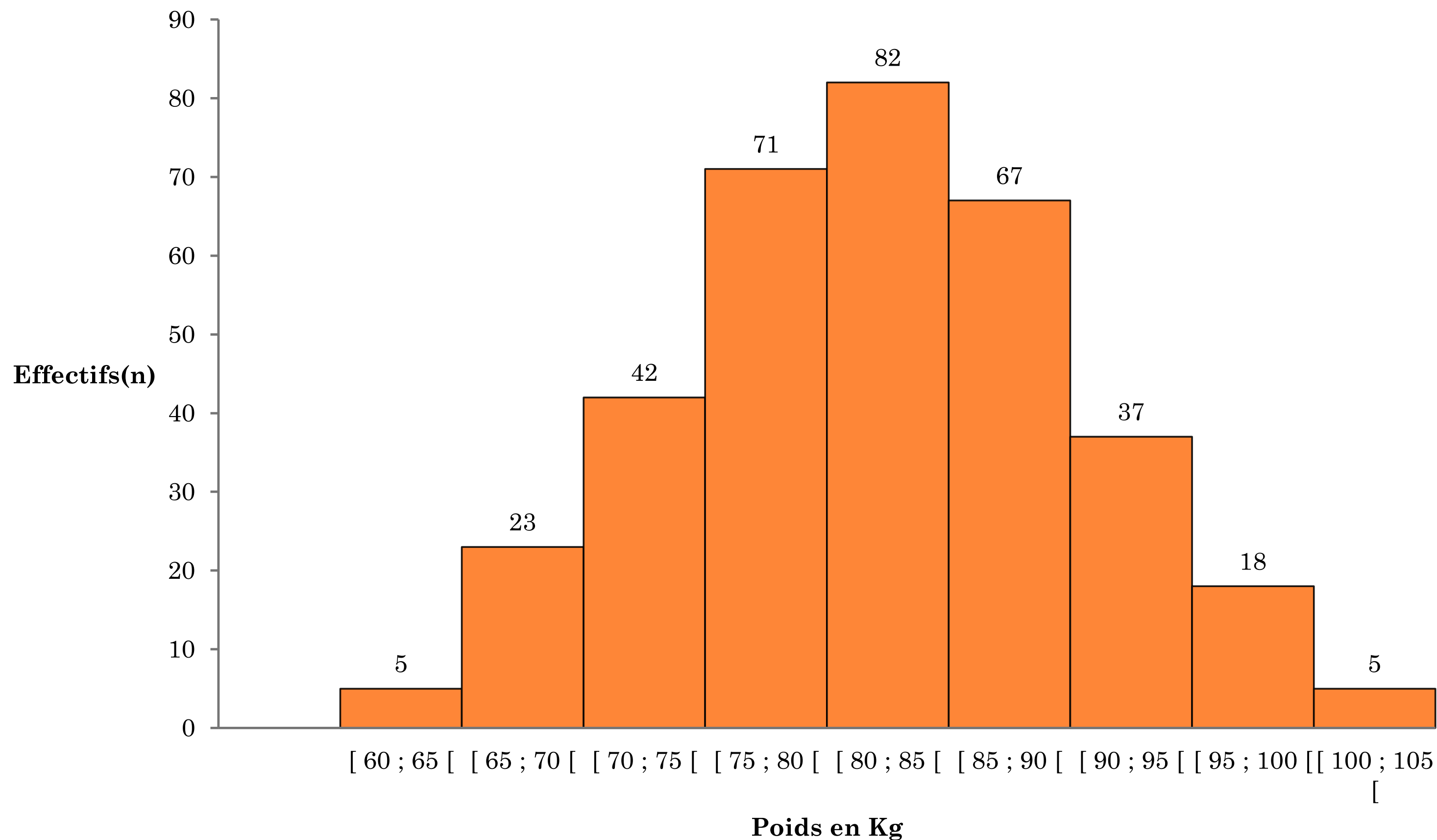


- Exercice 2: le tableau ci-dessous donne le poids de 350 personnes. Faire la représentation graphique correspondante.

Borne inférieure	Borne supérieure	Classes	Centres de classes	Effectifs
60	65	[60 ; 65 [	62,5	5
65	70	[65 ; 70 [	67,5	23
70	75	[70 ; 75 [	72,5	42
75	80	[75 ; 80 [	77,5	71
80	85	[80 ; 85 [	82,5	82
85	90	[85 ; 90 [	87,5	67
90	95	[90 ; 95 [	92,5	37
95	100	[95 ; 100 [	97,5	18
100	105	[100 ; 105 [	102,5	5
Total				**350**



HISTOGRAMME



Polygone de fréquences

polygone de fréquences

fréquence (n)

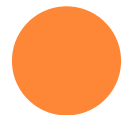
90
80
70
60
50
40
30
20
10
0

57,5 62,5 67,5 72,5 77,5 82,5 87,5 92,5 97,5 102,5 107,5

poids en Kg

Histogramme des effectifs

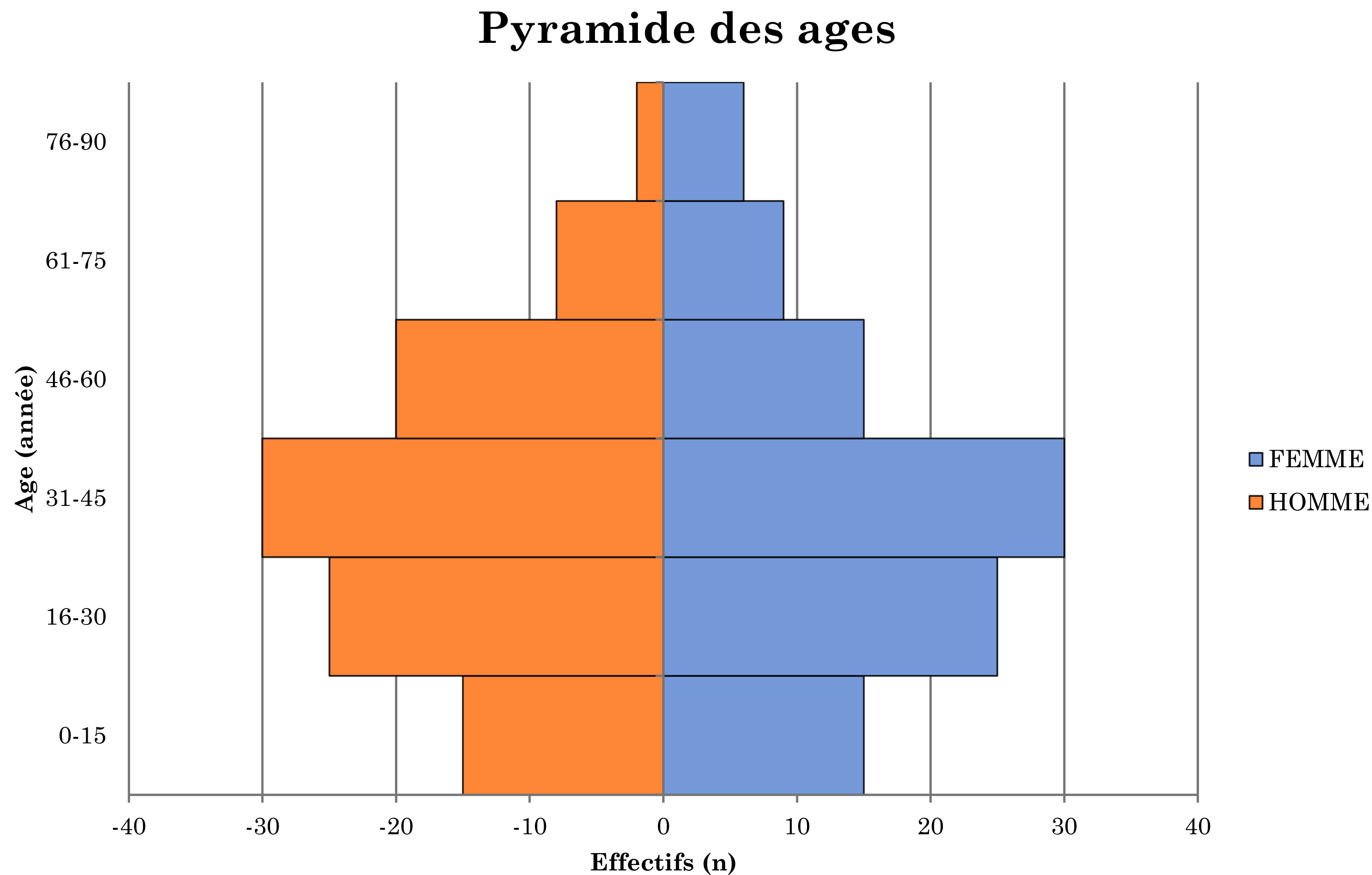
Polygone des effectifs



- Exercice 3: soit le tableau suivant

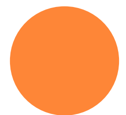
CLASSES AGE	HOMME	FEMME
0-15	-15	15
16-30	-25	25
31-45	-30	30
46-60	-20	15
61-75	-8	9
76-90	-2	6

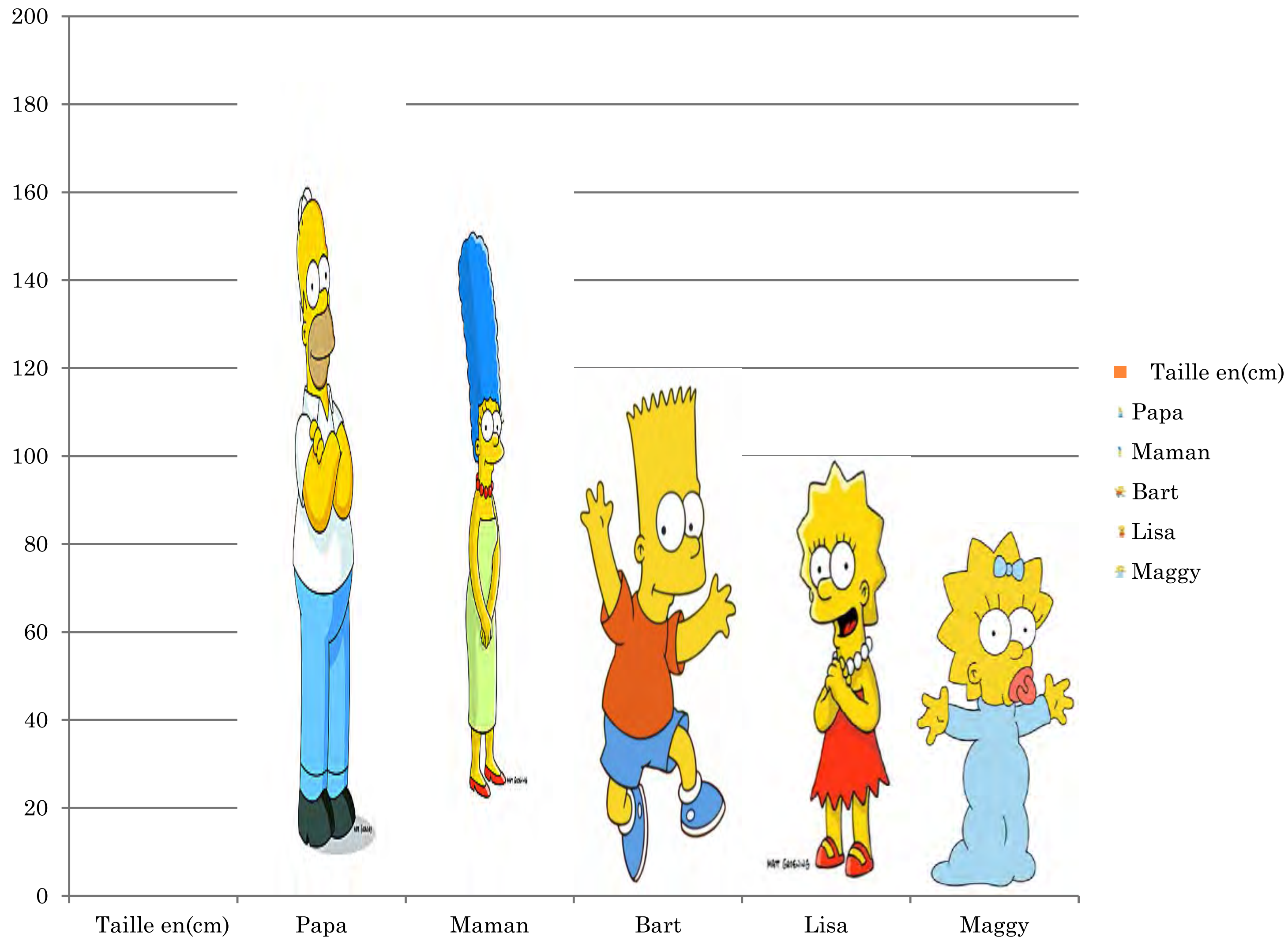
Pyramide des ages



Histogramme (on peut inclure des images..)

Taille en(cm)	
Papa	185
Maman	165
Bart	120
Lisa	100
Maggy	85




Histogramme:Distribution de la taille de 5 individus
200
180
160
140
120
100
80
60
40
20
0
Taille en(cm)
Papa
Maman
Bart
Lisa
Maggy
Taille en(cm)
Papa
Maman
Bart
Lisa
Maggy